Lexical Item Exchanges in the Noisy Channel Model: A replication and extension of Poppels and Levy, 2016

Rujul Gandhi

11 May 2019

Introduction

Communication through language is not error-free. Daily, we come across utterances that are obscured by some type of noise. This noise could be in the form of interruptions from the environment, phonological or syntactical changes from the speaker, or errors in the listener's comprehension. Thus, the utterances that humans convey to each other are often imperfect. Nevertheless, we manage to comprehend each other and respond appropriately, suggesting that we interpret language through a 'noise filter' that extracts the intended utterance from the perceived utterance. This idea is the basis of the noisy channel model of communication.

In 2013, Gibson, Bergen and Piantadosi proposed that humans are using a Bayesian optimization model when interpreting language through a noisy channel (Gibson et. al. 2013). According to this model, listeners will decide whether to interpret a perceived sentence A as an intended sentence B by weighing two factors: the probability that B would morph into A due to noise (Probability B → A), and the probability that B would be uttered in the first place (Probability (B)). Both of these factors are positively correlated to non-literal interpretation. This model can be applied to syntactic noise, specifically insertions and deletions of function words, which lead to sentences that are implausible according to world knowledge. For instance, deleting a single word changes the plausible sentence "The woman gave the candle to her daughter," to the implausible sentence "The woman gave the candle her daughter." The implausible sentences may then be interpreted literally, where the candle received something, or non-literally, where the daughter received the candle.

Using their Bayesian model, Gibson et. al. made a few predictions about the conditions under which implausible sentences would be interpreted non-literally with higher or lower probability. The findings show four key things:

- Sentences with insertions are interpreted non-literally less often than sentences with deletions. Since insertions are less likely, the probability that B → A is reduced.
- Sentences with more errors are interpreted non-literally less often than those with fewer errors (more errors reduce the probability that B → A)
- The filler materials influence non-literal interpretation. A higher number of implausible sentences (reducing probability that a plausible B is intended) will reduce non-literal interpretation, while a higher number of errors (increasing probability that B → A) will increase non-literal interpretation.

Gibson et. al.'s work provided ground for additional work in this field, specifically exploring other ways in which these baseline predictions could be verified or expanded. One extension, by Gibson, Tan, et. al. in 2017, investigates the effect of a speaker's accent on the listener's interpretation. The authors find that ubjects listening to American English sentences spoken with a non-native accent are more likely to interpret implausible sentences non-literally, compared to subjects listening to sentences spoken in a native accent. This experiment essentially deals with the probability of intended sentence B morphing into perceived sentence A. A listener naturally expects a higher probability of noise if they believe that the speaker is not native in the language. The Gibson, Tan, et. al extension builds upon Gibson et. al.'s 2013 baseline theory, adding both a piece of evidence to the basic model and a direction of study branching off of it.

While the 2013 study and consequent extensions classify syntactic noise into insertions or deletions, noisy utterances may also contain other types of string edits. The most obvious type is an exchange of lexical items, as below.

- The book fell from the *table* to the *floor.* (plausible)
- The book fell from the *floor* to the *table.* (implausible)

As seen above, exchange of the two noun positions turn a sentence from a plausible to an implausible one. One can argue that an exchange can be thought of as a sequence of deletions and insertions- deleting two lexical items from their original positions, and then inserting them at each other's original positions. This would lead to a much higher order edit than simple insertions or deletions, which would lead us to predict that if an exchange were being perceived as a sequence of insertions and deletions, the probability of one occurring is extremely low, and hence the probability of non-literal interpretation is very low or negligible. However, a study by Poppels and Levy in 2016 found that exchanges are also accounted for by the noise model: that is, a sentence like the implausbible one above is significantly more likely to be interpreted non-literally than a plausible one. Poppels and Levy's study, which looks specifically at exchanges of nouns in sentences involving prepositions, thus establishes that exchanges can be considered as a string edit of their own, rather than a sequence of other string edits.

In addition to studying exchanges as a type of noise, Poppels and Levy's study also introduces another independent variable: the effect of word order on perceived noise. Through corpus analysis, the authors establish that there exists a canonical order of appearance for most preposition pairs ('from...to', 'to...about', 'from...until', etc.) The opposite order is not ungrammatical, but it is improbable enough to be non-canonical. In other words, a phrase is much more likely to have a canonical word order such as 'from the X to the Y' than a non-canonical word order such as 'to the Y from the X'. Using a noncanonical word order may be compared to Gibson et. al.'s increase in filler material error rate. When a listener or reader is perceiving an implausible sentence, will they be more likely to interpret it non-literally if the word order is noncanonical as well? Gibson et. al.'s model suggests that this may happen. Accordingly, Poppels and Levy's study also shows a small but significant effect of canonicality on probability of literal interpretation. Thus, Poppels and Levy vary two independent variables, plausibility and canonicality, and find significant effects of both on non-literal interpretation when dealing with exchanges.

In the first of the experiments presented here, I have replicated Poppels and Levy's study of noun exchanges in English prepositional phrases. My primary goal for this replication is verifying the significance of the plausibility effect, and thus adding evidence to establish that exchanges are a separate type of string edit, which is accounted for by the human 'noise filter'. I also hope to verify that the correlation between word order canonicality and literal interpretation probability is significant, which would indicate that the functioning of the noise filter takes word order into account.

As an extension of the above, the second experiment presented in this paper studies the exchange of case markers in the Marathi language. Marathi is an Indo-Aryan language spoken in the Maharashtra region of India. There are two key reasons why a study of Marathi is interesting. First, it is a case-marking language, where nouns are modified by phonological markers to specify their roles in a sentence. Thus, exchanging the case markers on nouns changes a noun's syntactic role in a sentence. Secondly, word order in Marathi is more flexible than that of English, but the canonical word order is SOV. Alternating between SOV and OSV provides an analog to the canonicality variable in the Poppels and Levy study. Both these features are illustrated below.

- Case marker exchanges: alternating for plausibility

Structure in Marathi	**Translation**	
girl-SUB ball-OBJ kicked	The girl kicked the ball.	(plausible)
girl-OBJ ball-SUB kicked	The ball kicked the girl.	(implausible)

- Word order: alternating for canonicality

Structure in Marathi	**Translation**	
girl-SUB ball-OBJ kicked	The girl kicked the ball.	(canonical)
ball-OBJ girl-SUB kicked	The girl kicked the ball.	(noncanonical)

Building upon the hypotheses and results from Poppels and Levy, we expect that plausibility will affect non-literal interpretation, indicating that exchanges of case markers are also treated as a single string edit. We also expect that OSV word orders will have a higher rate of nonliteral interpretation than SOV word orders.

Methods

Replication

The replication experiment was in the form of a reading-comprehension task carried out through Amazon's Mechanical Turk platform. A total of 40 participants in the United States were asked to read sentences and provide answers to Yes/No questions displayed simultaneously with the sentences. The sentences were the same as in the materials from Poppels and Levy's original experiment. Sentences of interest contained two prepositional phrases and had four variations each, with plausibility and canonicality being the controlled, independent variables. For example, the sentence ”The package fell from the table to the floor,” is plausible and has canonical word order. Four variations of this sentence are possible:

- (+C, +P) The package fell from the table to the floor.
- (-C, +P) The package fell to the floor from the table.
- (+C, -P) The package fell from the floor to the table.
- (-C, -P) The package fell to the table from the floor.

Twenty such sentences, with four variations each, were used. Each sentence would be followed by a question, such as, ”Did the package fall from the floor?” The question was the same within each sentence, regardless of which of the four variations was displayed. Using the turkolizer algorithm, a unique order and combination of these twenty sentences along with sixty filler sentences was generated for each participant (i.e. each participant saw 80 sentences total). Only subjects who lived in the USA, indicated English was their native language, and answered over 75% of fillers correctly qualified for further analysis. After removing participants that did not meet one or more of these criteria, 35 participants remained.

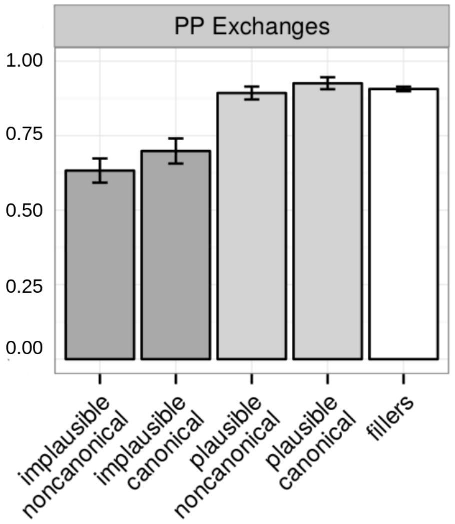

Figure 1: Percentage of literal interpretation, by category.

Extension

The extension was in the form of a reading-comprehension task carried out through Amazon's Mechanical Turk platform, with participants in India. The instructions and materials were written entirely in Marathi. Participants were asked to read sentences and provide answers to Yes/No questions displayed simultaneously with the sentences. Similar to the replication, sentences of interest had four variations each, with plausibility and canonicality being the controlled, independent variables. For example, for the sentence "The woman brought water from the river," the four variations would be as follows:

bai-ne nədi-tun paɳi aɳle	woman-SUB river-ABL water brought	(+C, +P)
nədi-tun bai-ne paɳi aɳle	river-ABL woman-SUB water brought	(-C, +P)
nədi-ne bai-tun paɳi aɳle	river-SUB woman-ABL water brought	(+C, -P)
bai-tun nədi-ne paɳi aɳle	river-SUB woman-ABL water brought	(-C, -P)

Twenty such sentences, with four variations each, were used. Each sentence would be followed by a question, such as, "Did the woman bring water?" The question was the same within each sentence, regardless of which of the four variations was displayed. Using the turkolizer algorithm, a unique order and combination of these twenty sentences along with thirty filler sentences was generated for each participant (i.e. each participant saw 50 sentences total). Only subjects who lived in India, indicated Marathi was their native language, and answered over 75% of fillers correctly qualified for further analysis. Of a total 84 instances of the task, participants were filtered according to this criteria and 11 participants remained.

Results

The results of Poppels and Levy's original experiment showed significant correlations between plausibility and literal interpretation as well as canonicality and literal interpretation. Their results are visually represented in figure 1. The white column is the base literal interpretation rate for filler materials. The light gray columns are literal interpretation rates for plausible utterances, while the dark gray columns are literal interpretation rates for implausible sentences. In both the plausible and implausible utterances, a small difference is seen in rates when canonicality is altered.

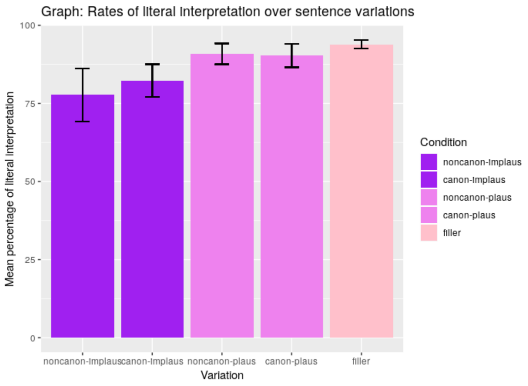

Figure 2: Replication experiment: percentage of literal interpretation for each category.

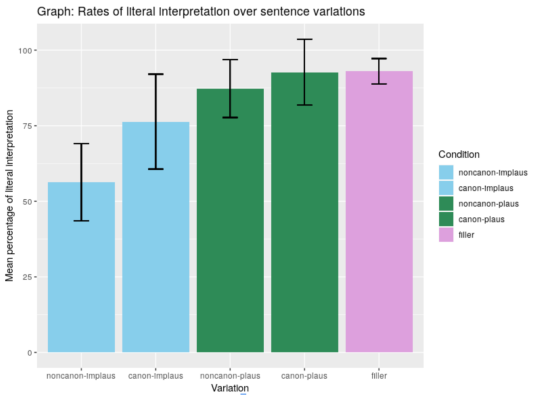

Figure 3: Extension with Marathi: percentage of literal interpretation for each category.

The graph shows that in Poppels and Levy's original experiment, plausible sentences were interpreted literally about as often as fillers, while implausible sentences were interpreted literally at a siginificantly lower rate. Each pair of gray columns can also be compared with each other to look at the effect of canonicality. Poppels and Levy found a significant main effect of both plausibility (= 1.303, $p < 0.0001$) as well as canonicality (= 0.276, $p = 0.038$) upon running a binomial mixed-effects regression analysis on their data.

In the replication, similar data was obtained. Figure 2 shows the percentage of literal interpretation by category, as obtained in the replication experiment. The light pink column corresponds to filler materials, which show a high (93.90476) percentage of literal interpretation. The plausible variations, represented by dark pink columns, show similar rates of literal interpretations both when they are canonical (90.28571) and non-canonical (90.85714). The implausible variations, corresponding to the columns in violet, show lower literal interpretations both for canonical (82.28571) and non-canonical (77.71429) forms.

After fitting the replication data to a binomial regression model, a significant effect of implausibility on literal interpretation (β = -0.43519, $p = 0.000126$) is found. However, there is no significant effect of canonicality ($p = 0.628272$). While the implausibility findings are in line with Poppels and Levy's original experiment, the canonicality findings are not.

The Marathi extension shows the same trends as Poppels and Levy's original as well as the replication with regards to plausibility. In figure 3, fillers are represented by the pink column, which shows a high level of literal interpretation. Green columns correspond to plausible sentences, which have similar levels of literal interpretation to the fillers. Blue columns correspond to implausible sentences, which are interpreted literally with significantly less probability.

On fitting the data to a binomial regression model, a significant effect of plausibility on literal interpretation ($p < 0.05$) is found, as well as a significant effect of canonicality on literal interpretation ($p < 0.05$).

Discussion

The correlation of plausibility and literal interpretation, specifically as translated to exchanges of lexical items, was replicated in this experiment as well as in the extension. This means that implausible sentences were more likely to be interpreted non-literally than plausible ones. The implications of this for the noisy channel model are that it gives evidence for exchanges as a type of noise which the human filter is equipped to deal with. It establishes exchanges as a separate type of edit, and not a sequence of insertions and deletions which is less likely to occur- since they are still expected to some degree by the human noise filter and sentences with exchanges are treated pragmatically. Additionally, exchanges of case markers in Marathi still followed the trends of exchanges in Poppels and Levy. Even though switching of case markers impacted the syntactic roles of nouns, readers were still interpreting these sentences pragmatically. The idea of exchanges as separate and valid noise types thus can translate beyond English, to a case-marking language.

Interestingly, the effect of canonicality on literal interpretation did not appear in the replication experiment, but it did replicate in the extension experiment. This suggests that canonicality of word order may or may not have a significant effect on whether sentences are interpreted literally, and further experimentation is required to come to a conclusion. It is interesting to note that Poppels and Levy, while establishing the canonical word orders of different preposition pairs, found some pairs more strongly favoring a certain order than others. For instance, ”from... until” was favored over ”until...from” in nearly all instances, but ”for...in” and ”in...for” were nearly equal in their occurrences. Pairs that do not lean strongly in a canonical direction may be causing this disrepancy in measuring canonicality effects. The judgement of canonicality is non-binary and may depend upon subjects’ idiolects. Another possible explanation for the disrepancy is that people may be interpreting the prepositions not separately, but as part of the larger prepositional phrase which includes the noun- for instance, ’from the table’. The prepositions in general may have more or less frequently used word orders. However, if people perceive the nouns as being moved around, since nouns do not have any canonical ordering, this may lead to a disregard for canonicality.

It is also possible that limitations of the replication experiment led to missing the significance of canonicality. A smaller number of participants was a key difference between the replication and orginal experiments. While the original experiment had 59 valid participants, the replication had 35. The original experiment, however, was also testing multiple types of edits and alternations, while the replication was focused entirely on prepositional phrase exchanges. The filler materials used in the replication may also have had some effect on the results. Some fillers were also implausible, as they had been taken from Poppels and Levy’s replication of Gibson et. al. 2013.

Effects of canonicality may be stronger in the Marathi extension since canonicality here deals with the basic word order (SOV vs OSV) of the sentence. A non-canonical word order, as discussed previously, may be acting in a manner similar to the non-native accent of Gibson, Tan, et al 2017, by priming the reader to have expectations about how much noise is probable. A corpus analysis of Marathi texts may help establish a more concrete measure of canonicality and help build a hypothesis about whether or not the canonical constructions used in my materials are more canonical for Marathi than the constructions in Poppels and Levy are for English.

The extension in Marathi was largely limited by the online format of the experiment. Although participants were screened for multiple criteria, the experimental set-up could not be as controlled as in a physical setting. A majority of participants had to be removed, and it was hard to ascertain participants' understanding of the task as opposed to guessing or confusion. This led to a small sample size. Hence, I propose to replicate the Marathi experiment with native speakers in a physical setting. I hope to get a larger sample size, of 40-60 speakers, and have a controlled set-up. I may also be able to collect demographic data and consider results from that angle.

As my experiment was not limited geographically to the Marathi-speaking region within India, there is a chance that a mixture of L1 and L2 speakers were obtained. L1 and L2 speakers of a language may, however, deal with noise in different ways. It may be possible to do an extension that looks at whether these differences exist, and how they manifest if they do. This extension would ideally also be in the form of a physical, in-person task rather than an online task.

Finally, all three of the experiments discussed in this paper- Poppels and Levy 2016, the replication, and extension with Marathi- deal with 2 controlled variables and four resulting variations of each sentence. Thus, the effects of plausibility and canonicality may be interacting with each other. It would be interesting to separate these variables and look at each one entirely on its own. For instance, if a case-marking language with a relatively rigid word order was studied, the variable of canonicality could be removed entirely and only the effects of plausibility with case-marker exchanges would be seen. This would help cement the hypothesis of lexical item exchanges as a valid noise type and the ability of the noisy channel model to interpret them.

References

1. Poppels, T. and Levy, R. P. (2016). Structure-sensitive Noise Inference: Comprehenders Expect Exchange Errors.

2. Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences, USA, 110, 8051–8056.*

3. Gibson, E., Tan, C., et. al. (2017). Don't Underestimate the Benefits of Being Misunderstood, *Psychological Science*